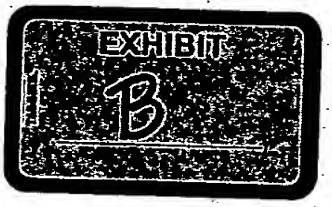
INTERNATIONAL ORGANISATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
MOVING PICTURES AND ASSOCIATED AUDIO

CODING OF



ISO/IEC JTC1/SC29/WG11
MPEG96/1084
July 1996, Tampere

Title: Results of Scalability Experiments

Source: R. L. Schmidt, A. Puri and B. G. Haskell (AT&T)

Status: Proposal

1. Introduction

At the previous MPEG meeting, bidirectional video object planes (B-VOPs) based coding was added to the MPEG-4 Video VM. B-VOPs based coding combines the flexibility of MPEG-1 B-pictures and bits savings efficiency of H.263 PB-frame (B-blocks) by allowing the choice of prediction in forward, backward, bidirectional and the direct modes, on a macroblock basis. Furthermore, B-VOPs, similar to I- and P-VOPs can be of arbitrary shape. B-VOPs, since they are noncausal (do not feedback into the interframe coding loop) they can be easily separated to enable Temporal scalability, which is one of the very efficient forms of scalability. The core experiment B1 includes rectangular VOPs based Temporal Scalability while the core experiment C1 includes arbitrary shaped VOPs based Temporal Scalability.

In this document, we present results of core experiment B1 using the B-VOP prediction modes of the current VM. Further optimization of these modes may be possible for increased coding efficiency by using the prediction modes originally suggested in B1 (MPEG96/N1266) and is expected to be investigated for September meeting. The proposed syntax for VM including the scalability syntax is described in MPEG96/1047. A document from Sharp Corp. (MPEG96/1043) provides results of experiment C1 which was found to be very closely related and necessitated development of the proposed combined syntax. Due to the fact that video objects (VO's) can be both completely arbitrary or rectangular in shape, experiments B1 and C1 are considered to be two parts of the same experiment rather than two separate experiments.

2. Temporal Scalability with B-VOPs

Scalability involves use of two or more layers, for the purpose of experiments we are considering two two layer scalability only. The two layers are, a base-layer and an enhancement-layer. As is typical, the base-layer refers to the layer coded indpendently whereas the enhancement-layer is the layer coded dependently with respect to the base-layer. In addition to inter-frame prediction that is employed in nonscalable (single layer) coding, typical forms of scalability such as Temporal

scalability and Spatial scalability also use inter-layer prediction. Normally, one of the main differences between Spatial and Temporal scalability is that Spatial scalability does not use motion vectors for inter-layer predictin due to its temporal coincidence but Temporal scalability does.

As mentioned earlier, our experiment is on Temporal scalability with rectangular VO's and B-VOP coding. All of the B-VOPs belong to enhancement layer whereas base-layer carries I- and P-VOPs. In general, this is not a requirement for the proposed scalability, in fact, the MPEG-2 syntax on which our proposed scalability syntax is based allows use of I-, P- or B-pictures in the base-layer. In our experiments, the prediction structure employed for the base- and the enhancement-layers is shown in Figure 1.

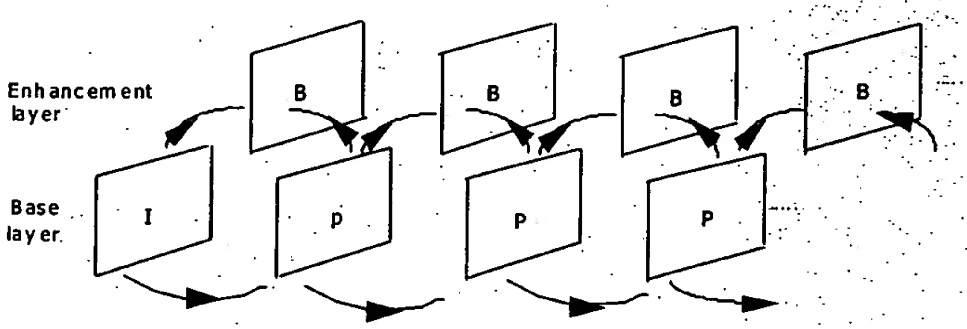


Figure 1 Prediction Structure employed in two layer B-VOP based Temporal scalability

Since, these experiments are intended for low bitrates in MPEG-4, the original video sequences at CIF or QCIF resolutions are temporally decimated by a factor of 2 before coding begins. Thus, the prediction structure of Figure 1 applies to the temporally decimated input sequence and not to the original input sequence, and further this temporally decimated sequence is demultiplexed to form two sequences, one input to the base-layer encoder and other to the enhancement-layer encoder. Our base-layer encoder is same as the VM2.2 encoder with or without B-VOPs, whereas the enhnacement-layer encoder uses B-VOPs only and interlayer predictions with respect to decoded base-layer VOPs.

3. Experiment Results

We now present the results of our experiments on Temporal scalability using I- and P-VOPs in the base layer and B-VOPs only in the enhancement layer. Two test conditions are employed, first, QCIF sequences coded at a total of 24 kbits/s, and second, CIF sequences coded at a total of 112 kbit/s. In both case, the frame rate of the base-layer is 5 frame/s and that for the enhancement layer is 10 frames/s; temporally multiplexing of base and enhancement layers result in 15 frames/s. The results of our experiments are shown in Table 1 for QCIF resolution and in Table 2 for CIF resolution.

Table 1 Results of Temporal scalability experiments with QCIF sequences at a total of about 24 kbits/s

Sequence	Layer and frame rate	VOP type	QP	SNR Y dB	SNR Cb dB	SNR Cr dB	Avg. Bits per VOP	Avg. Bitrate kbits/s
Akiyo	Enh. layer @ 10 frames/s	B	20	32.24	34.09	36.55	991	9.91
	Base layer @ 5 frames/s	I/P	14.14	32.33	34.14	36.54	1715	8.58
Silent	Enh. layer @ 10 frames.s	B	25	28.73	33.77	35.44	1452	14.52
	Base layer @ 5 frames/s	I/P	19.02	28.93	33.80	35.50	2614	13.07

Mother & Daughter	Enh. layer @ 10 frames/s	B	19	32.79	38.20	38.89	1223	12.23
	Base layer @ 5 frames/s	I/P I	12.08	33.05	38.32	39.00	2389	11.95
Container	Enh. layer @ 10 frames/s	B	20	30.02	36.65	35.75	1138	11.38
	Base layer @ 5 frames/s	I/P	14.14	30.06	36.64	35.74	2985	14.93

Table 2 Results of Temporal scalability experiments with CIF sequences at a total of about 112 kbits/s

Sequence	Layer and frame rate	VOP type	QP	SNR Y	SNR Cb	SNR Cr	Avg. Bits	Avg. Bitrate kbits/s
Akiyo	Enh. layer @ 10 frames/s	B	27	32.85	35.13	37.88	4821	48.21
	Base layer @ 5 frames/s	I/P	22.1	32.85	35.13	37.88	5899	29.50
Mother & Daughter	Enh. layer @ 10 frames/s	B	27	32.91	37.71	38.12	6633	66.33
	Base layer @ 5 frames/s	I/P	22.1	32.88	37.86	38.30	8359	41.80
Silent	Enh. layer @ 10 frames.s	B	29	29.07	34.35	35.83	5985	59.85
	Base layer @ 10 frames/s	I/P	24.06	29.14	34.36	35.89	9084	45.42.
Container	Enh. layer @ 10 frames/s	B	29	28.47	36.17	35.50	6094	60.94
	Base layer @ 5 frames/s	I/P	24.06	28.52	36.21	35.55	8325	41.63
News	Enh. layer @ 10 frames/s	B	29	29.84	34.04	35.78	6197	61.97
	Base layer @ 5 frames/s	I/P	23.08	29.70	34.04	35.76	9710	48.55
Coastguard	Enh. layer @ 10 frames/s	B	29	26.64	37.71	40.86	10739	107.39
	Base layer @ 5 frames/s	I/P	24.06	27.04	37.83	41.00	14938	74.69
Foreman	Enh. layer @ 10 frames/s	B	29	29.82	36.38	36.87	9864	98.64

	Base layer @ 5 frames/s	I/P	24.12	30.00	36.42	36.87	14459	72.29

4. Summary

With in the context of core experiment B1, we have experimented with Temporal scalability at low bit-rates using the MPEG-4 VM tools and the new proposed syntax. In our experiments, rectangular VOPs (frames) were used. The base-layer consisted of an I-VOP followed by P-VOPs, whereas, the enhancement-layer consisted of B-VOPs only. One-third of the total coded frames belong to the base-layer, the remaining two-third belong to the enhancement-layer. Total bitrate is nearly equally split between the base- and enhancement-layers.

Results in Table 1 and 2 verify the performance of the proposed solution for Temporal scalability for MPEG-4 applications at low bitrates; this solution is based on recently introduced B-VOPs and thus the performance of B-VOPs based coding is also verified. In addition, the proposed reorganized syntax structure which includes scalability was successfully used in this experiment (as well as in experiment C1) and is thus also verified. Further, some optimization of prediction modes for B-VOPs, in context of scalability or otherwise is possible. The proposed syntax for scalability also enables Spatial scalability of rectangular or irregular shape VO's as a specific mode in generalized scalability.